JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2002, **78,** 433–447 NUMBER 3 (NOVEMBER)

CONTEXTUAL CONTROL OF STIMULUS GENERALIZATION AND STIMULUS EQUIVALENCE IN HIERARCHICAL CATEGORIZATION

KAREN GRIFFEE AND MICHAEL J. DOUGHER

CONCORD COLLEGE AND UNIVERSITY OF NEW MEXICO

The purpose of this study was to determine whether hierarchical categorization would result from a combination of contextually controlled conditional discrimination training, stimulus generalization, and stimulus equivalence. First, differential selection responses to a specific stimulus feature were brought under contextual control. This contextual control was hierarchical in that stimuli at the top of the hierarchy all evoked one response, whereas those at the bottom each evoked different responses. The evocative functions of these stimuli generalized in predictable ways along a dimension of physical similarity. Then, these functions were indirectly acquired by a set of nonsense syllables that were related via transitivity relations to the originally trained stimuli. These nonsense syllables effectively served as names for the different stimulus classes within each level of the hierarchy.

Key words: categorization, hierarchical categorization, stimulus generalization, stimulus equivalence

Categorization has been extensively investigated in the psychological literature. Categories are generally defined as sets of individual stimuli that share stimulus functions, sometimes but not always on the basis of physical similarity (e.g., Anglin, 1977; Flavell, 1985; Lea, 1984). That is, stimuli may share formal properties, as in the category ''red things,'' or they may share functions, as in the category ''things I'll work to get, like ice cream, money, or approval.'' As has been pointed out by others (e.g., Adams, Fields, & Verhave, 1993; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields, Reeve, Adams, & Verhave, 1991), natural categories consist of stimuli that share both physical features and stimulus functions.

A special case of categorization is hierarchical categorization, or the creation of subcategories within categories. Developmental psychologists have typically approached the study of categorization, including hierarchical categorization, descriptively. Typical research questions include the age at which children can perform categorization tasks or the kind of categorization typical for children at different ages. For example, hierarchical categorization is typically studied through the use of class-inclusion logic problems. A child may be presented with 20 pieces of candy, 10 of which are M&Ms® and 10 of which are jelly beans. The experimenter then asks the child whether there are more M&Ms or more candies. It has been found that 5- and 6-year-olds will tend to compare M&Ms and jelly beans rather than M&Ms and the total number of candies, as asked. Seven- and 8-year-olds, however, are likely to perform the task correctly (McCabe, Seigal, Spence, & Wilkinson, 1982).

Developmental psychologists, especially those within the Piagetian tradition, explain these age-based results of the ''jelly bean challenge'' by describing qualitative changes in thinking that are believed to occur with development. It is believed that thinking is determined by schemes, which are ''like intellectual computer software that directs and determines how data from the world are looked at and dealt with'' (Achenbach, 2001, p. 157). Young children are said to have ''centration,'' a cognitive set that causes them to focus their attention on the most salient physical features of the stimuli (Piaget, 1929).

From a behavior-analytic or functional-analytic perspective, this research suggests that the categorizing behavior of younger chil-

This research was presented at the annual meeting of the Association for Behavior Analysis, May 1997, in Washington, D.C. Prior to that it was the doctoral dissertation of the first author, who never would have completed or published it without the support and friendship of the former ''Black Scorpion Society/Dougher Lab'' members at the University of New Mexico. We also thank the reviewers for their contributions, particularly Yvonne Barnes-Holmes for her thoughtful additions. Scott Rehorn was also an invaluable contributor via his computer program and pilot work.

Direct correspondence and requests for reprints to Karen Griffee, Department of Psychology, Concord College, Athens, West Virginia 24712 (e-mail: griffeek@concord.edu).

dren is frequently under a particular kind of stimulus control, namely the formal characteristics of the stimuli. By the age of 7 or 8 years, the source of stimulus control shifts so that children tend to make verbal categorization responses on the basis of stimulus functions. That is, although M&Ms and jelly beans *are* different in appearance, the children's answers to the logic problem reflect their experience with contingencies that support differential responses to the two types of candies. These contingencies are likely under contextual control. In some contexts, the two kinds of candies occasion different response classes. In other contexts, the candies occasion the same responses. That is, children learn that in some contexts distinguishing between M&Ms and jelly beans is irrelevant. For example, eating either before dinner will result in punishment if ''candy'' has been forbidden. It seems likely that having a candy-specific learning history that supports responding appropriately in a variety of contexts enables a correct response to the ''jelly bean challenge.''

The purpose of the present study was to address this issue experimentally within a behavior-analytic framework. In particular, we attempted to produce hierarchical categorization in a laboratory setting. If the idea of ''hierarchy'' is useful at all, it is in reference to the kinds of contingencies that control and maintain the behavior of categorizing. Some contingencies support responding to physically different stimuli as if they are the ''same,'' whereas other contingencies support responding to physically similar stimuli as if they are ''different.'' Stimuli involved in naturalistic hierarchical categories often have some physical features in common but are generally categorized according to function. For example, a domestic cat may be correctly categorized (labeled) as a pet, a feline, a mammal, and an animal, as we go from the specific level to the general level of the animal hierarchy. Our verbal categories may reflect real aspects of the physical characteristics of that organism. We can, however, also explain our use of those terms by noting that there are conditions under which it is most useful and appropriate to respond to the cat very generally as an animal, more specifically as a mammal, or still more specifically as Aunt Edna's beloved ''Fluffy.'' In each context, our behavior is determined by somewhat different behavioral contingencies.

The first part of this experiment asked whether hierarchical functional classes of physically related stimuli can be established in a laboratory setting. Specifically, we wanted to determine whether it was possible to establish the kind of hierarchical categories typically referred to in everyday language. These typical categories seem to be composed of members that share a stimulus function. Furthermore, this function may covary with the members' physical similarity. For example, cats of various fur lengths can be treated as if they are the same (i.e., as if they are simply cats rather than dogs). In some contexts, however, it may be necessary to further distinguish between long- and short-haired breeds, and only to a lesser extent continue to categorize them all as ''cats.''

In this experiment, these ''cats'' were represented by a series of triangle stimuli that varied along a continuous physical dimension. Color contexts were used to represent the circumstances under which it would be appropriate to treat all the cats as if they are the same or to distinguish among them. At the top of our hierarchy (green context), all triangles function similarly. In the middle of our hierarchy (red context), the triangles represent functionally distinct short- and long-haired breeds. At the bottom of our hierarchy (yellow context), the triangles represent four different breeds ranging from very short to very long hair.

An additional feature of typical hierarchical categories is that they are unidirectionally inclusive. That is, Fluffy is a cat, and all cats are mammals, and all mammals are animals; but the reverse is not true. All animals are not mammals, all mammals are not cats, and all cats are not ''Fluffy.'' In this experiment, we attempted to establish a specific form of contextual control over class membership such that the general function shared by all of the triangle ''cats'' at the top of the hierarchy would transfer to the triangle ''cats'' at the lower levels, but that the reverse would not occur. In this sense, the transfer of functions within the hierarchical category would be unidirectional.

The second part of this experiment asked whether arbitrary stimuli (nonsense syllables) can be brought into the categories and whether these stimuli would acquire the functions of names or labels. According to the developmental literature, preverbal infants first learn functional categories and only later label them (e.g., Bee, 1999; Feldman, 1999). This labeling is important because it allows the almost infinite expansion of the category in the absence of any direct training or experience with individual members. Once a stimulus is labeled as belonging to an existing category, it automatically acquires the functions shared by the members of that category. According to Feldman (1999), critical cognitive and behavioral advances are thought to develop along with the acquisition of ''symbolic function,'' or the ability to use a ''mental symbol,'' that is, a word or object representing something not physically present. For example, ''preschoolers can use a mental symbol for a car (the word car), and they likewise understand that a small toy car is representative of the real thing. Because of their ability to use symbolic functions, children have no need to get behind the wheel of an actual car to understand its basic purpose and use'' (p. 192). Furthermore, symbolic function also allows children to think beyond the present to the future and to consider multiple possibilities at the same time.

The processes by which verbal stimuli are brought into already existing functional classes and the transfer of functions that results from these processes are central to the stimulus equivalence literature. This literature has most recently been absorbed by and interpreted within relational frame theory (summarized in Hayes et al., 2001). Stimulus equivalence is a phenomenon in which functional relations among stimuli emerge without direct training when a human is taught a series of interrelated conditional discriminations (Sidman & Tailby, 1982). For example, if a child learns through conditional discrimination training to match A to B and A to C, it is likely that he or she will, without additional training, be able to do more than this. Specifically, he she will be likely to show *reflexivity* by matching each of these stimuli to itself; *symmetry* by matching B to A and C to A; and *transitivity* by matching B to C and C to B. The stimuli involved in the conditional discrimination become, in this context, functionally substitutable for each other. Therefore, they are said to be *equivalent,* or members of an equivalence class (Sidman, 1971; Sidman & Tailby, 1982). In this way, there is what could be called an explosion of knowledge that occurs in an emergent way. With the direct training of only 15 relations, for example, 60 untrained conditional relations may emerge (Sidman, Kirk, & Willson-Morris, 1985). A symbol and its referents can also be seen to form an equivalence class. For example, Sidman (1971) asked a young retarded man to select pictures in response to their spoken names (A-B). The subject was then taught to select the pictures' corresponding written names in response to their spoken names (A-C). Without further training, the subject then matched the written names to the pictures (B-C) and vice versa (C-B), and also named the pictures (B-A) and read the printed words (C-A).

Hayes et al. (2001) have argued that equivalence relations are examples of arbitrary relations among stimuli, because they are not based on the formal characteristics of the stimuli but emerge through the training of interrelated conditional discriminations. In other words, these relations are arbitrary because behaving as if the stimuli are equivalent is established entirely by reinforcement. For example, there is no formal relation between the word *dog* and a dog, but the practices of the social-verbal community have taught us to behave in some ways as if they are the same. It has been demonstrated that when an equivalence class is trained through conditional discrimination training, the stimuli may then also share stimulus functions, or become *functionally equivalent* in some ways (e.g., Dougher, Augustson, Markham, & Greenway, 1994; Wulfert & Hayes, 1988). When a child learns to tact a shared category function, such as ''animals you pet,'' or to label a category, such as ''cats,'' this necessarily involves stimulus equivalence because the relation between the category label and the referent is arbitrary. It has been hypothesized by Sidman (1989) that ''the most important function of equivalence relations is to transfer new stimuli—for example, words—into already existing functional classes'' (p. 273). Accordingly, the procedures used to establish stimulus equivalence classes were used in the present study to bring nonsense syllables into existing functional classes. Thus, it was predicted that the contextual control of stimulus generaliza-

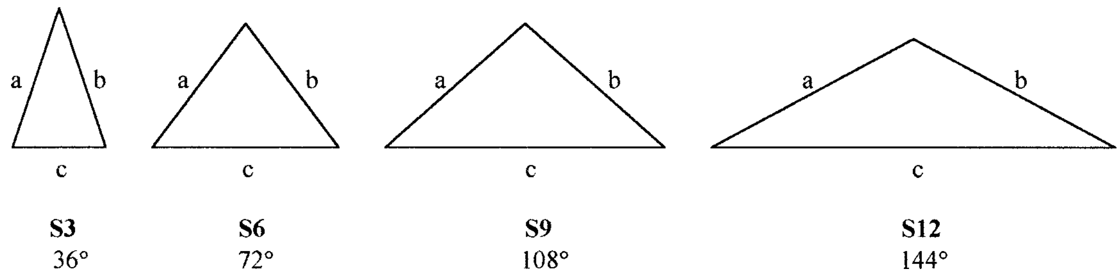

Fig. 1. Stimuli used in training.

tion and stimulus equivalence would establish hierarchical categorization in a laboratory setting.

METHOD

Subjects

Introductory psychology students were recruited for participation through in-class announcements. The first 5 students to volunteer served as subjects. They received course credit for their participation and also earned money dependent on their performance. At the beginning of the experiment, the general procedures were explained, and all subjects read and signed a statement of informed consent. Subjects were free to discontinue at any time. All subjects were debriefed after completion of the study.

Setting, Materials, and Apparatus

Each subject worked individually in one of two small (approximately 2 m by 2.5 m) rooms. Each room contained a chair, desk, keyboard, monitor, and personal computer, which was used to present stimuli and record subject responses. The desk was situated along the far wall facing away from the doorway, so the subject could be intermittently observed with minimal distraction.

Stimuli were designed to enhance the issue of physical similarity as a relevant dimension. These stimuli were 14 triangles (designated S1, S2, . . ., S14) that varied in terms of the angle formed by Sides A and B and, correspondingly, the length of Side C (see Figure 1). Sides A and B were always 1.5 in. long, and the triangles were positioned so that Side C was the horizontal base. Otherwise all triangles were the same. They were drawn with a black line and appeared on a red, green, or yellow background. The AB angles ranged from 12° to 168° and varied in increments of 12°. Thus, for each stimulus, the AB angles were as follows: S1 = 12, S2 = 24, S3 = 36, S4 = 48, S5 = 60, S6 = 72, S7 = 84, S8 = 96, S9 = 108, S10 = 120, S11 = 132, S12 = 144, S13 = 156, S14 = 168. Other stimuli consisted of the seven nonsense syllables bup, yap, wug, zig, pif, gak, and git.

Procedure

The study consisted of two parts, each with a training and test phase. The training phase of Part 1 consisted of a series of conditional discriminations presented in different background colors on the computer screen. The training goal was to bring responding to Angle AB/Side C length of four sample triangles under the contextual control of the background colors (Figure 2). Responses consisted of selecting a button from an array presented along the bottom of the computer screen. The correct button depended on the triangle and the background color presented on a given trial. The test phase of Part 1 entailed the presentation of novel triangles along with the training stimuli, and was intended to assess whether differential responding generalized to these novel stimuli. In testing, trials were presented in extinction.

The training phase of Part 2 was similar to that of Part 1 in that it also consisted of a discrimination task intended to bring differential responding to four sample triangles under contextual control. Responses in this task, however, consisted of matching nonsense syllables to the triangles. The test phase of Part 2 consisted of three different tests. Test 1 determined whether naming responses would generalize to novel triangles. Test 2 determined whether the original stimulus functions trained in Part 1 would transfer to the

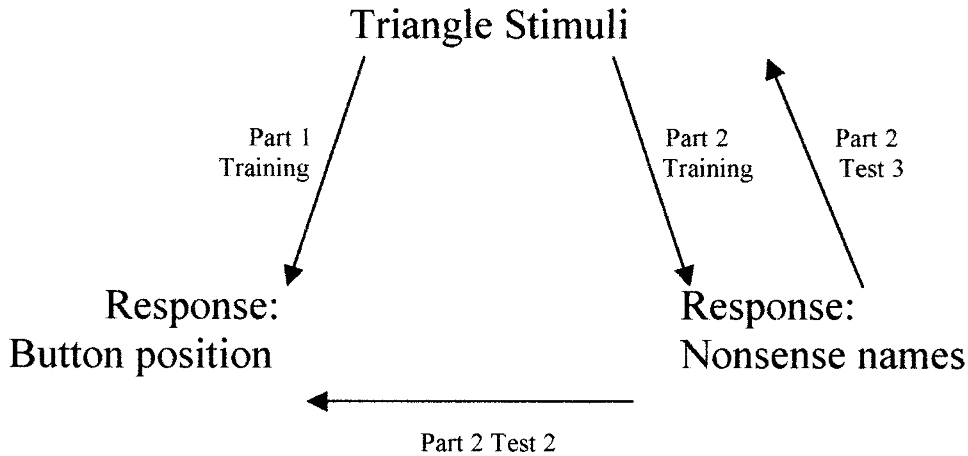

Fig. 2. Graphic depiction of the trained and tested relations among stimuli used in this experiment.

syllables. Test 3 assessed whether symmetrical relations between the nonsense syllables and novel triangles would emerge.

Part 1: Conditional Discrimination Training of Selection Responses

During this phase, subjects received the following instructions:

> On the computer screen, you will see a background color, a triangle at the top of the screen, and seven buttons at the bottom of the screen. You earn points by choosing the correct button, depending on what you see on the screen. It might take a while to learn which is the correct button, because the triangles are all very similar. But, there will always be a "best" answer, which is worth 6 points. Sometimes there will also be a "next-best" answer, worth 3 points, and a "next-next-best" answer, worth 1 point. The rest of the answers will be wrong. Your job is to earn as many points as possible! You will be able to cash in your points for money, up to $20.00. We will also keep track of who earns the most points of all the subjects this semester. That champion will win $40.00.

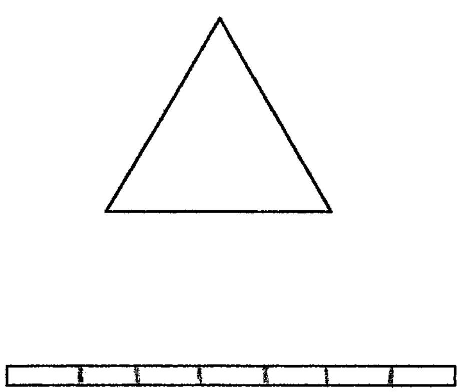

Fig. 3. A sample trial as seen by the subject during conditional discrimination training of selection responses.

The training phase involved a conditional discrimination task in which subjects were taught to respond differently to four training triangles, depending on color background. Subjects were alerted to attend to background color simply to facilitate training. One of four sample triangles (S3, S6, S9, and S12) appeared on the top-center of the computer screen, followed immediately by the presentation of seven buttons across the bottom of the screen (Figure 3).

Subjects responded by using a mouse to click on one of the buttons. Responses were coded by the button (left to right) that the subjects selected: R1 indicates the selection of Button 1, R2 indicates Button 2, and so on. After a button was selected, the screen cleared and written feedback appeared on the screen. Depending on a subject's response, the feedback presented was "6 points. This is the best answer"; "3 points. But there is a better answer"; "1 point. But there are better answers"; or "Incorrect. This is a wrong answer."

Which response was best depended jointly on the color background of the computer screen and the sample (see Figure 4). When the background color was green, subjects earned six points for selecting the middle

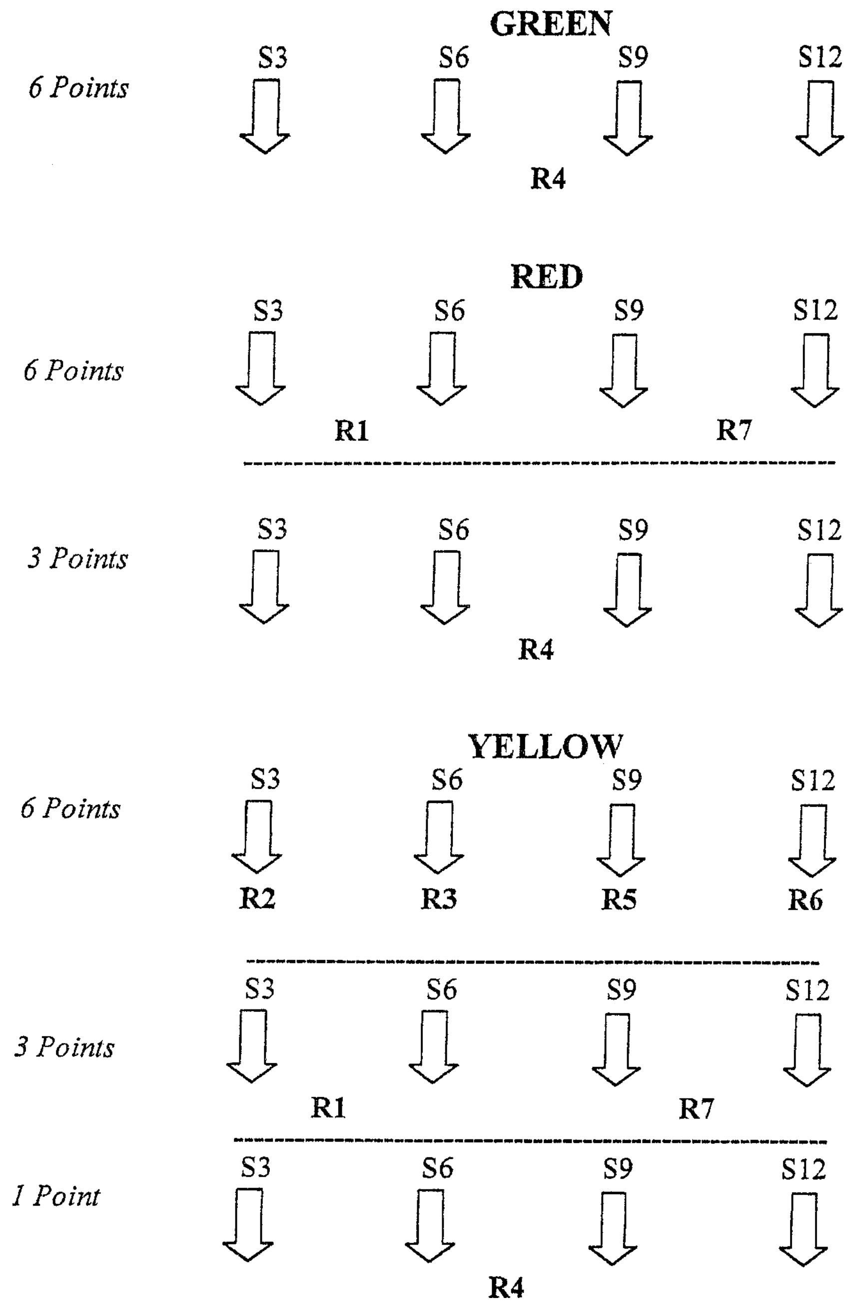

Fig. 4. Part 1: conditional discrimination training of selection responses. Points earned for button position responses (R1 through R7) to sample triangles (S3, S6, S9, S12) by background color.

button (i.e., R4) in the presence of any of the four training stimuli. All other responses were incorrect. When the background color was red, subjects earned six points for selecting R1 in the presence of S3 and S6 and for selecting R7 in the presence of S9 and S12. If they selected R4 in the red background, they earned only three points. All other responses were incorrect. In the yellow background, the six-point responses were R2 in the presence of S3, R3 in the presence of S6, R5 in the presence of S9, and R6 in the presence of S12. Three points were awarded for R1 in the presence of S3 or S6 and R7 in the presence of S9 or S12. One point was awarded for R4 in the presence of any of the stimuli.

Training trials began in the green background. One of the four sample stimuli was presented individually in quasirandom order until the subject reached the training criterion described below. After the criterion was met in the green background, the sample stimuli were presented in quasirandom order in the red background until the criterion was met. Then the trials previously presented in the green and red backgrounds were randomly mixed and presented again until the subject again met the training criteria. After the training criterion was met with these mixed trials, training trials were presented in the yellow background. When criterion was reached in the yellow background, a final mixed task was presented in which sample stimuli were presented in randomly alternating red, green, and yellow backgrounds until the subject met criterion. The training criterion for each background color on this final task was the same as for each individual color condition.

The training criterion in the green background was 95% correct responses after a minimum of 50 trials. The maximum number of trials required to meet this criterion was 110. No three-point or one-point responses were possible in the green background. The training criteria in the red background required subjects to choose the six-point response at least 90% of the time and to choose a three-point response 5% or less after a minimum of 50 trials. The maximum number of trials required to meet this criterion was 178. No one-point responses were possible in the red background. To satisfy training criteria in the yellow background, subjects were required to choose the six-point response at least 88% of the time, a three-point response 5% or less, and a one-point response 2% or less after a minimum of 50 trials. The maximum number of trials required to meet this criterion was 200. On the final mixed task during the training phase, the above criteria were required for each color context for a minimum of 150 trials. The maximum number of trials required to meet this criterion was 320. When criterion was met on this final task, the computer screen cleared and instructions for the test phase appeared.

Part 1: Generalization Test

Instructions for the test phase were as follows:

> This next part will be a little different from the last. Even though correct answers still earn the same points, you won't get feedback after your choices: we won't tell you how many points you're earning anymore. You will also probably see some new triangles during this part. Even though there will be new triangles and you won't see the points you earn, there is always a best, 6-point answer and sometimes other answers are worth 3 points or 1 point. Pay attention to the background color when you choose the buttons and try to earn as many points as possible to earn money and win the championship.

Testing consisted of a series of trials presented without feedback to determine whether the trained responses generalized to novel triangles. The triangles included all of the triangles described previously, including the four training triangles depicted in Figure 1 and the remaining 10 novel triangles that were not presented during training. Each of the 14 test stimuli was quasirandomly presented five times in each of the three background colors, for a total of 210 test trials.

Part 2: Conditional Discrimination Training of Naming Responses

The instructions, procedures, and training criteria in this part were identical to the training phase of Part 1 with one exception. The subjects learned a naming response rather than a button-selection response in the presence of the sample triangles. Seven, nonsense syllables (wug, yap, zig, git, bup, gak, and pif) appeared across the bottom of the screen on each trial. Subjects were instructed to match particular names to sample stimuli by selecting a name in the same way they selected a button in Part 1, except that the names appeared in randomly alternating positions along the bottom of the screen on each trial.

Which response was the best answer again depended jointly on the color background and sample triangle (see Figure 5). When the background color was green, subjects earned six points for selecting wug in the presence of any of the four training stimuli. All other

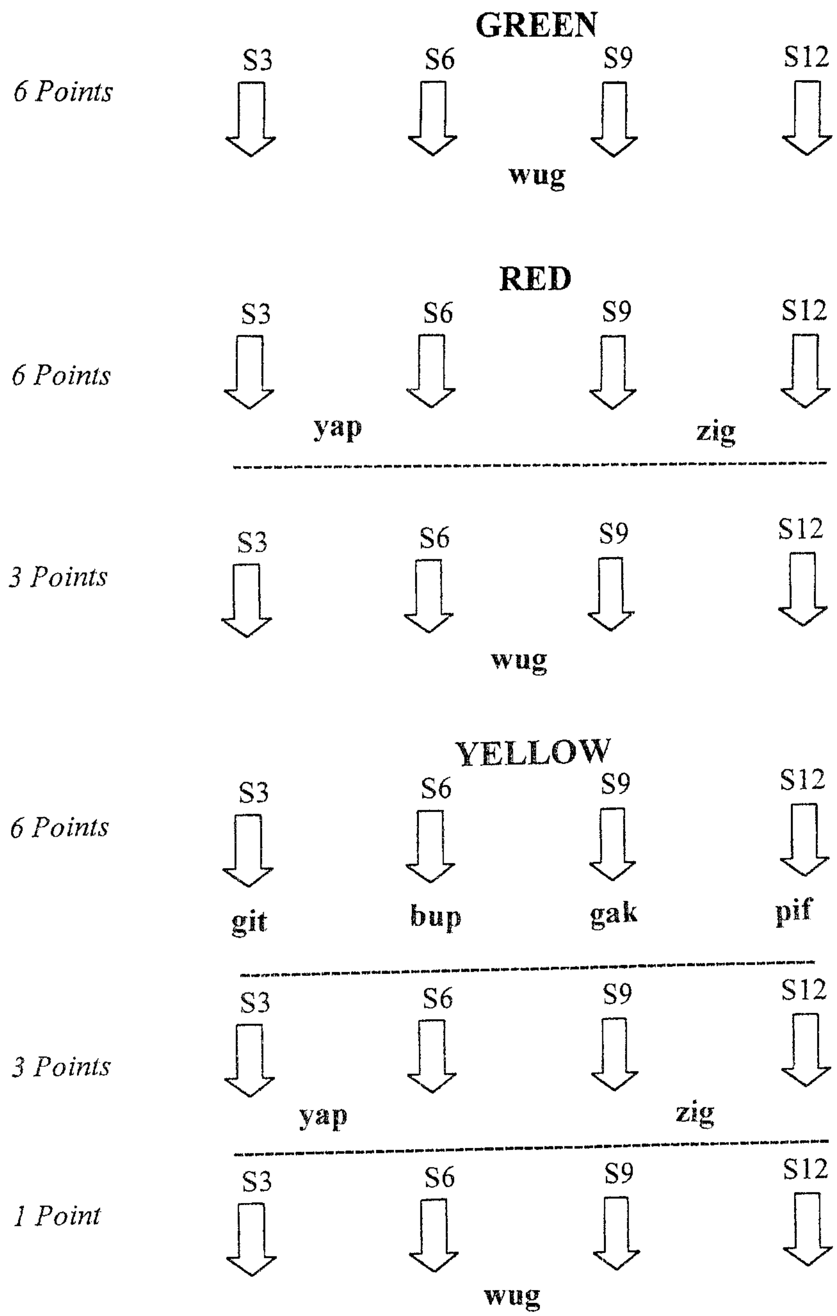

Fig. 5. Part 2: conditional discrimination training of naming responses. Points earned for nonsense-word matching to sample triangle stimuli (S3, S6, S9, S12) by background color.

responses were incorrect. When the background color was red, subjects earned six points for selecting yap in the presence of S3 or S6 and for selecting zig in the presence of S9 or S12. Selections of wug earned three points. All other responses were incorrect. In the yellow background subjects earned six points for selecting git in the presence of S3, bup in the presence of S6, gak in the presence of S9, and pif in the presence of S12.

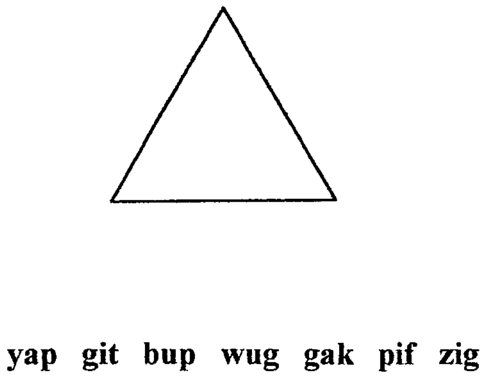

Fig. 6. A sample trial as seen by the subject during conditional discrimination training of naming responses.

Subjects earned three points for selecting yap in the presence of S3 and S6 and for selecting zig in the presence of S9 and S12. Selections of wug earned only one point in the yellow background.

Part 2: Generalized Naming, Transitivity, and Generalized Symmetry Tests

There were three tests in the test phase of Part 2. The generalized naming test was a replication of the generalization test of Part 1, except that subjects selected nonsense syllables rather than buttons (see Figure 6). The instructions and trial presentations were identical to those in Part 1. The purpose of these trials was to determine whether the trained naming responses would generalize to the novel triangles.

When the generalized naming test was completed, the screen cleared and the instructions for the transitivity test appeared. Subjects were again told that correct answers earned the same points, but they would not receive feedback. The transitivity test consisted of a series of trials presented without feedback to determine whether the original stimulus functions trained in Part 1 (i.e., the selection responses) would transfer to the nonsense syllables. Subjects previously learned (a) to select buttons in the presence of triangles and (b) to select nonsense names in the presence of the same triangles. The transitivity test determined whether, on the basis of this past training alone, subjects would then be able to select the correct buttons in the presence of the nonsense words and background colors.

The nonsense names used as response choices during the conditional discrimination training of naming responses appeared as samples at the top of the screen during the transitivity test. Each nonsense name was quasirandomly presented five times in the red, green, and yellow backgrounds for a total of 105 test trials. When this test was completed, the screen cleared and instructions for the generalized symmetry test appeared. Subjects received similar instructions, except that they were told the following:

> What's different about this next part is that we want you to choose whether or not the things you see on the screen go together. Choose a "YES" or "NO" button the same way you chose the other buttons earlier, to indicate whether or not the triangle and the letters you see go together.

The generalized symmetry test consisted of a series of trials presented without feedback to determine whether untrained symmetrical relations would emerge between the nonsense syllables and the novel triangles. These relations could be said to have emerged if subjects successfully matched triangles to the nonsense names, having been taught to match the names to the training triangles during conditional discrimination training (see Figure 2).

The triangles used in the generalized symmetry test were the same 14 used in the generalization test in Part 1. On each trial, a triangle appeared in the top center of the screen. It was immediately followed by the presentation of one of the nonsense names directly below the triangle. At the bottom of the screen, there also appeared a "yes" and a "no" button, from which the subjects were asked to choose. There was a total of 1,470 test trials comprised of five presentations of every possible combination of each of the 14 triangles with each of the seven nonsense names in each of the three background colors.

After this test, subjects were debriefed, given the appropriate participation points for class credit, and dismissed. The subjects were contacted and paid after the study was completed.

Triangle Stimuli

Subjects	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Green														
1	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
2	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
3	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
4	4-5	4-4,2-1	4-4,2-1	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-4,5-1	4-5	4-5	4-4,7-1
5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
Red														
1	1-5	1-5	1-5	1-5	1-5	1-5	1-4,4-1	1-2,7-3	7-5	7-5	7-5	7-5	7-5	7-5
2	1-5	1-5	1-5	1-5	1-5	1-4,3-1	1-4,4-1	1-1,7-4	7-5	7-5	7-5	7-5	7-5	7-5
3	1-5	1-5	1-5	1-5	1-4,3-1	1-3,4-2	1-1,7-4	1-1,7-4	7-5	7-4,1-1	7-5	7-5	7-5	7-5
4	1-5	1-5	1-5	1-4,3-1	1-5	1-5	1-5	1-2,7-3	7-5	7-4,5-1	7-5	7-4,6-1	7-5	7-5
5	1-5	1-4,4-1	1-5	1-5	1-5	1-5	1-2,7-3	7-5	7-5	7-5	7-5	7-5	7-5	7-4,4-1
Yellow														
1	2-5	2-5	2-5	2-4,3-1	3-5	3-4,2-1	3-5	5-3,3-2	5-5	5-3,6-2	6-4,5-1	6-5	6-5	6-5
2	2-5	2-5	2-5	2-4,3-1	3-5	3-5	3-5	5-5	5-5	5-3,6-2	6-4,5-1	6-5	6-5	6-5
3	2-5	2-5	2-5	3-5	3-3,4-2	3-3,4-2	4-5	5-3,4-2	5-4,4-1	5-3,6-2	6-5	6-4,5-1	6-3,7-2	6-1,7-4
4	2-5	2-5	2-5	2-1,3-4	3-5	3-5	3-5	5-3,4-2	5-3,4-2	5-5	6-3,5-2	6-4,5-1	6-5	6-5
5	2-3,1-2	2-5	2-4,3-1	3-5	3-5	3-4,7-1	3-2,5-3	5-5	5-5	5-5	6-1,5-4	6-5	6-5	6-4,7-1

Fig. 7. Part 1, generalization test data for all subjects. Selection responses to each triangle stimulus by background color. The first number in each cell indicates button position selected (1 through 7 possible). The second number indicates the times subjects made that selection response (1 through 5 possible; 0 responses are not graphed).

RESULTS

Part 1: Generalization Test

The data for the generalization test are presented in Figure 7. Subjects correctly selected buttons dependent on color context, and these responses generalized to the novel test stimuli. Because each test stimulus was presented five times in each background color, there are five possible responses in each cell.

Recall that the training stimuli were triangles S3, S6, S9, and S12. In the green background, subjects were trained to select Button 4 in the presence of each training stimulus. In testing in the green context, all subjects chose Button 4 in the presence of the novel stimuli as well. With the exception of Subject 4, who made a few selection errors (in 4 of 70 trials), Button 4 was selected in the presence of every stimulus by every subject.

Recall that in the red background subjects were trained to select Button 1 in the presence of S3 and S6 and to select Button 7 in the presence of S9 and S12. Button 4 also counted toward criterion up to 5% of the time. Again, these responses generalized to the novel stimuli for all subjects. As the test triangles approached the halfway point between S6 and S9, subjects became more likely to divide their responses between Buttons 1 and 7 or to default by selecting Button 4 (note responses to S6 through S9 for all subjects in the red background).

In the yellow background, subjects were trained to select Button 2 in the presence of S3, Button 3 in the presence of S6, Button 5 in the presence of S9, and Button 6 in the presence of S12. Responses appropriate to the red and green backgrounds were also acceptable 5% and 2% of the time, respectively. All subjects showed generalization to the novel stimuli. As in the red background, intermediate test stimuli were more likely to elicit mixed responding.

Part 2: Generalization Test

Results for the generalized naming test are presented in Figure 8. These results are almost identical to those of the generalization test in Part 1. This pattern of responding was

Triangle Stimuli

Background	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Green															
	2	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
	3	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
	4	w4,y1	w 5	w 5	w 5	w5, t1p1	w 5	w 5	w 5	w 4,k1	w 5	w 5	w 5	w 5	w 5
	5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
Red															
	1	y5	y5	y5	y5	y5	y 4, w1	y 5	y3, w1z1	z 5	z 5	z 5	z 5	z4, k1	z 5
	2	y5	y4,t1	y5	y5	y5	y 4	z 5	y3	z3, w1k1	z 5	z 5	z 5	z4, k1	z 5
	3	y5	y5	y5	y3, w2	y5	y 4	y2, z3	y3	z 5	z 5	z 5	z 5	z 5	z 5
	4	y5	y4,w1	y5	y5	y5	y4,w1	z3,w1p1	z2,w2y1	z3,y2	z4,b1	z 5	z 5	z4, k1	
	5	y5	y 5	y4, w1	y5	y5	y4, p1	z3, p2	z 5	z 5	z 5	z 5	z 5	z 5	z 5
Yellow															
	1	t 5	t 5	t 5	b3, t2	b4, t1	b5	b5	k3, b2	k 5	k 5	k4, p1	p 5	k4, p1	p 5
	2	t 5	t 5	t 5	t 5	b 5	b 5	b 5	b 5	k 5	k 5	p3, k2	p 5	p4, k1	p 5
	3	t 5	t 5	t 5	b4	b5,t1	b 5	b 5	k4, b1	k 5	k 5	k 5	p 5	p 5	p 5
	4	t3, w1y1	t3, w1z1	b2, z3	b3, t2	b5	b4, p1	k2, b1t2	p4, y1	k3, b2	k3, b2	k4, b1	p2, k2b1	p3, y2	p4, t1
	5	t 5	t 5	t 5	t 5	b3,t2	b4, w1	b4, k1	k 5	k 5	k 5	k 5	p 5	p 5	p3, w2

Fig. 8. Part 2: generalized naming test data for all subjects. Nonsense-syllable responses for each test triangle (S1 through S14) by background color. Each cell indicates the number of times subjects selected a particular syllable. Subjects could select any syllable for a total of five trials for each triangle stimulus. Only those syllables chosen are listed. Trigrams are listed by first letter, except gak (k) and git (t).

expected, given that the only procedural change from the training phases of Part 1 to Part 2 was that the response was naming rather than selection of a button.

Most subjects correctly selected nonsense names dependent on color, and those responses generalized to the novel test stimuli. Intermediate stimuli were again more likely to elicit mixed responding in the red and yellow backgrounds. Although Subject 4 reached criterion during the testing phases (as did all subjects), the trained responses in the yellow background dropped out during the nonreinforced test trials, and there was substantial variability in her pattern of responding.

Part 2: Transitivity Test

Recall that subjects first learned button selection responses in the presence of the triangles (Part 1) and then naming responses in the presence of the same triangles (Part 2). It was expected that subjects would then choose the appropriate buttons in the presence of nonsense names, given that this represents a transitive relation if stimulus equivalence classes formed. The data in Figure 9 show that the nonsense syllables became discriminative for choosing the expected buttons.

In the green background, first note the cells representing the Button 4 response in the presence of wug. Subjects previously learned to select wug (Part 2) and were trained to choose Button 4 (Part 1) in the presence of all triangles. In the transitivity test with the green background, all subjects chose Button 4 when wug was the sample stimulus. In the red background, subjects previously learned to select yap and Button 1 in the presence of S3 and S6; they also learned to select zig and Button 7 in the presence of S9 and S12. For all subjects in the transitivity test with the red background, yap became discriminative for choosing Button 1 and zig became discriminative for choosing Button 7. There was one exception, Subject 4, who chose Button 7 in the presence of yap once. In the yellow background, git was discriminative for a Button 2 response, bup was discriminative for Button 3, gak was discriminative for Button 5, and pif was discriminative for Button 6.

Transfer of these stimulus functions to the nonsense words was shown without error for all subjects except Subject 4. As previously noted, the naming responses originally trained in the yellow background dropped off during the nonreinforced test trials for this subject. Predictably, these functions trans-

Nonsense Syllables

Subjects	yap	git	bup	wug	gak	pif	zig
Green							
1	1-2,4-3	2-4,4-1	3-3,4-2	4-5	4-1,5-4	6-3,4-2	7-4,4-1
2	1-5	2-5	3-2,4-3	4-5	5-2,4-3	6-4,4-1	7-5
3	1-5	2-4,4-1	3-2,4-3	4-5	5-4,4-1	6-4,4-1	7-2,4-3
4	1-5	2-5	3-3,4-1,2-1	4-5	4-5	6-4,4-1	7-4,2-1
5	1-5	2-4,3-1	3-5	4-5	5-5	6-3,7-2	7-5
Red							
1	1-5	2-4,1-1	3-5	4-5	5-3,4-1,7-1	6-4,7-1	7-5
2	1-5	2-4,1-1	3-4,4-1	4-5	5-5	6-2,4-1,5-2	7-5
3	1-5	2-3,1-1,4-1	3-3,4-2	4-5	5-2,7-3	6-3,4-1,7-1	7-5
4	1-4,7-1	2-4,4-1	3-3,1-2	4-5	5-4,1-1	6-4,7-1	7-5
5	1-4,3-1	2-5	3-3,6-1,7-1	4-5	5-5	6-5	7-5
Yellow							
1	1-3,4-2	2-5	3-5	4-5	5-5	6-5	7-3,4-1,1-1
2	1-5	2-5	3-4,5-1	4-5	5-5	6-3,5-1,2-1	7-5
3	1-2,4-3	2-5	3-5	4-5	5-5	6-5	7-3,4-2,1-11
4	1-3,4-1,6-1	2-2,1-2,5-1	3-2,2-1,5-1,7-1	4-5	5-3,4-1,3-1	6-3,3-1,7-1	7-2,5-1,3-1,2-1
5	1-5	2-5	3-5	4-5	5-5	6-5	7-5

Fig. 9. Part 2: transitivity test data for all subjects. Selection responses to each nonsense syllable by background color. The first number in each cell indicates button position selected (1 through 7 possible). The second number indicates the times subjects made that selection response (1 through 5 possible; 0 responses are not graphed).

ferred poorly to the nonsense words in the yellow background during the transitivity test.

Although the relations between the nonsense names and button responses described above were the focus of the transitivity test, appropriate experimental control required that test trials include presentations of all nonsense words in all color backgrounds. These control presentations introduced a dilemma for the subjects in that these trials forced subjects to make some button responses to nonsense words that appeared in novel color backgrounds. For example, in the generalized naming test, wug was correctly matched to triangle stimuli only in the green background. In the transitivity test, wug also appeared in the red and yellow backgrounds, and subjects were required to make some button responses on those trials. On these trials, the button responses were expected to be either random, under the discriminative control of the color alone, or under the discriminative control of the nonsense name alone (via transitive stimulus equivalence relations, as noted previously). Note (except for Subject 4) the consistent pattern of responding on these control trials across subjects. If responses on these trials were random, no such clear pattern would be apparent, except on the test trials of primary interest described above. In contrast, responses on these control trials might predictably have come primarily under control of the color backgrounds, given the reinforcement history provided in Part 1. Had this been the case, the following response pattern would have been shown: green: Button 4 only; red: Buttons 1 and 7 only; yellow: Buttons 2, 3, 5, and 6. Subjects tended to err by giving these responses. For example, see Subject 1, transitivity test, green background (Figure 9) and note that Button 4 was chosen 10 of 20 times during control (non-wug) trials. However, the results show that nearly all button responses by all subjects during this phase were controlled by the nonsense words, regardless of color. That is, the discriminative functions for the button choice transferred to the nonsense names even in the trials with novel color–name combinations.

Part 2: Generalized Symmetry Test

The generalized symmetry test demonstrated the emergence of untrained relations sym-

Triangle Stimuli

Subjects	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Green														
1	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
2	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
3	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
4	w 5	w 5	w 4	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 4	w 5	w 5	w 5
5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5	w 5
Red														
1	y5	y5	y5	y5	y 5	y 5	y 2,z3	y5	z 5	z 5	z 5	z 4	z 5	z 4
2	y5	y5	y5	y5	y5	y 5	y 3, z4	y 1,z4	z 5	z 5	z 5	z 4	z 5	z 4
3	y5	y5	y5	y5, w1	y5	y 5	y 2, z5	z 5	z 5	z 5	z 5	z 5	z 5	z 5
4	y5	y 5	y5	y5	y4	z4, y2w1	z3,pwyt1	z 5, b1	z 5	z 5	z 4	z 5	z 4	z 5
5	y5	y 5, p1	y 5, p2	y5	y3	y 5, z2	y 1, z3	y 5, z5	z 5	z 5	z 5	z 5	z 5	z 5
Yellow														
1	t 4	t 5	t 5,b3	t 3, b5	b5	b5	b 5, k2	b 5, k5	k 4, p2	k 4, p5	p 5	p 5	p 5	p 5
2	t 5	t 5	t 5, b5	t 2, b5	b 5	b 5	b 5, k2	b 5, k5	k 5, k4	p 5, k4	p 5	p 5	p 4	p 5
3	t 5	t 5	t 5	t3, b5	b 5	b 5	b 5	k 5, b2	k 5	k 5, p1	k 5	p 5	p 5	p 5
4	t 1	t 3, b2	t 4, b1	b 2	b 4	k1	k 2, b1	k1	k2	k 3,p1	k 2, p1	p 3, k3	p 3, k1	p 4
5	t 2	t 5	t 5	t 3, b3	b 4	b 5, k1	k 4, b2	k 5, b5	k 4	k 3, p2	k 5	p 5	p 4	p 1

Fig. 10. Part 2: generalized symmetry test data for all subjects. ''Yes'' matches between nonsense syllables and each test triangle (S1 through S14) by background color. A maximum of five ''yes'' responses is possible in each cell. Fewer than five ''yes'' responses in each cell necessarily means that the remaining responses were ''no'' responses. Trigrams are listed by first letter, except gak (k) and git (t).

metrical to those learned during conditional discrimination training of naming responses. Specifically, subjects matched the appropriate training triangles to the nonsense syllables and the matching responses also generalized to the novel triangles. Figure 10 illustrates these results by color context for each subject.

Recall that the instructions for this test phase asked subjects to make ''yes'' or ''no'' judgments about whether the test triangles and nonsense names went together. Data presented for the generalized symmetry test are the number of ''yes'' responses to each name–triangle pair. Because each pair was presented five times, a maximum of five ''yes'' responses is possible in each cell. Fewer than five ''yes'' responses in each cell necessarily means that the remaining responses were ''no'' responses.

All subjects showed symmetrical responding in each background color. This symmetrical responding also generalized to the novel test stimuli.

DISCUSSION

This study produced hierarchical categorization through the contextual control of functional classes established via the interrelated processes of conditional discrimination and generalization. Recall that in the training conditions of Parts 1 and 2, subjects demonstrated differential responses (button selection and a naming response) to the triangles depending on color background. Thus, differential responding to a specific stimulus feature was brought under contextual control. This contextual control was also hierarchical in that the stimuli at the top of the hierarchy had a more general stimulus function than those at the lower levels. That is, the stimuli at the top of the hierarchy comprised one large functional class, whereas the stimuli at the bottom formed several smaller functional classes. As was expected, the function shared by all of the stimuli at the top of the hierarchy was occasionally shared or acquired by stimuli at the lower levels. This sharing or transfer of functions among the stimuli in the hierarchy, however, was unidirectional in that stimulus functions acquired at the bottom levels were rarely acquired by the stimuli at the top. When a bottom-up transfer did occur, it most often involved novel triangles with angles halfway between those of the training triangles. For example, subjects presented with stimuli in the green background almost never erred by giving responses appropriate to the red and yellow backgrounds. Subjects presented with novel triangles in the red and yellow backgrounds, however, did occasionally default by giving responses appropriate in the green background, particularly when the angles of the triangles were halfway between those of the training triangles. An analogy for these results would be the following: Suppose a camper knows the difference between a mountain lion and a bear. She suddenly sees a large, furry creature that looks strangely like both. Her immediate response (leaving) is likely to reflect the shared stimulus functions of bears and mountain lions. If she verbally identifies the creature, she might default in her answer by saying, ''Well, I just know it was a big mammal!''

Similar analogies will be described to suggest the relevance of the other results of this experiment. In Part 1, subjects matched button selection responses to triangles, depending on background color. This might be analogous to learning that there are specific contexts in which it is appropriate to respond to several cats as individual cats, whereas in other, more general contexts, it is more appropriate to respond to all the cats as if they were alike. In the generalized naming test in Part 2, subjects were able to match the triangles to nonsense syllables by context. This might be analogous to learning that in specific contexts, we should refer to the cats by their individual pet names; in the general context, we can refer to them all as ''animals.'' Once this learning has occurred, it is suggested that the names normally become redundant with the context: If we are given care instructions that refer to a cat by its pet name rather than as an ''animal,'' it usually follows that the context in which we are expected to respond is one appropriate to treating that animal as a pet.

The present results also demonstrated that hierarchical functional classes could be extended by bringing arbitrarily related verbal stimuli into the classes. The transitivity test of Part 2 demonstrated that the originally trained functions (button choices) within the hierarchy transferred to the nonsense syllables, so that the syllables became names for the response classes within each level in the hierarchy. Presumably, these untrained responses occurred via the emergence of transitive equivalence relations. The generalized symmetry test of Part 2 then showed that the degree to which the triangles shared a particular function also seemed to be the degree to which subjects would name them as class members by matching them to the nonsense stimuli. These findings demonstrate that arbitrary stimuli such as words can come to stand for categories of stimuli when the categories are hierarchically arranged. These results support Sidman's (1989) notion that an important function of stimulus equivalence is to bring arbitrary stimuli such as words into preexisting functional classes.

A possible limitation of this study is that the subjects were adults with preexisting well-developed repertoires relevant to categorizing and naming. As a result, it is not possible to conclude that the experimental procedures were solely responsible for the behavior that emerged or that these kinds of procedures are responsible for emergence of hierarchical categorization in natural settings. Nevertheless, the results do support the contention that the present procedures were sufficient to produce hierarchical categorization in verbally competent subjects, and they are consistent with the assertion that stimulus generalization and equivalence play a role in the development of hierarchical categorization. As such, the study achieved its primary goal.

REFERENCES

Achenbach, T. A. (2001). Developmental psychopathology. In R. S. Feldman (Ed.), *Child development* (2nd ed., pp. 155–187). Upper Saddle River, NJ: Prentice Hall.

Adams, B., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record, 43,* 553–567.

Anglin, J. M. (1977). *Word, object, and conceptual development.* New York: Norton.

Bee, H. L. (1999). *Lifespan development* (2nd ed.). Boston: Addison-Wesley.

Dougher, M. J., Augustson, E., Markham, M. R., & Greenway, E. E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62,* 331–351.

Feldman, R. S. (1999). *Child development: A topical approach.* Upper Saddle River, NJ: Prentice Hall.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68,* 67–91.

Fields, L., Reeve, K., Adams, B. J., & Verhave, T. (1991). The generalization of equivalence relations: A model for natural categories. *Journal of the Experimental Analysis of Behavior, 55,* 305–312.

Flavell, J. (1985). *Cognitive development* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of language and cognition* (pp. 21–49). New York: Plenum.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 263–276). Hillsdale, NJ: Erlbaum.

McCabe, A. E., Siegel, L., Spence, L., & Wilkinson, A. (1982). Class-inclusion reasoning: Patterns of performance from three to eight years. *Child Development, 53,* 780–785.

Piaget, J. (1929). *The child's conception of the world.* New York: Harcourt, Brace.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., & Tailby, W. (1982). Conditional discrimination versus matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Wulfert, E., & Hayes, S. C. (1988). The transfer of conditional sequencing through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125–144.

Received November 28, 2001
Final acceptance June 17, 2002